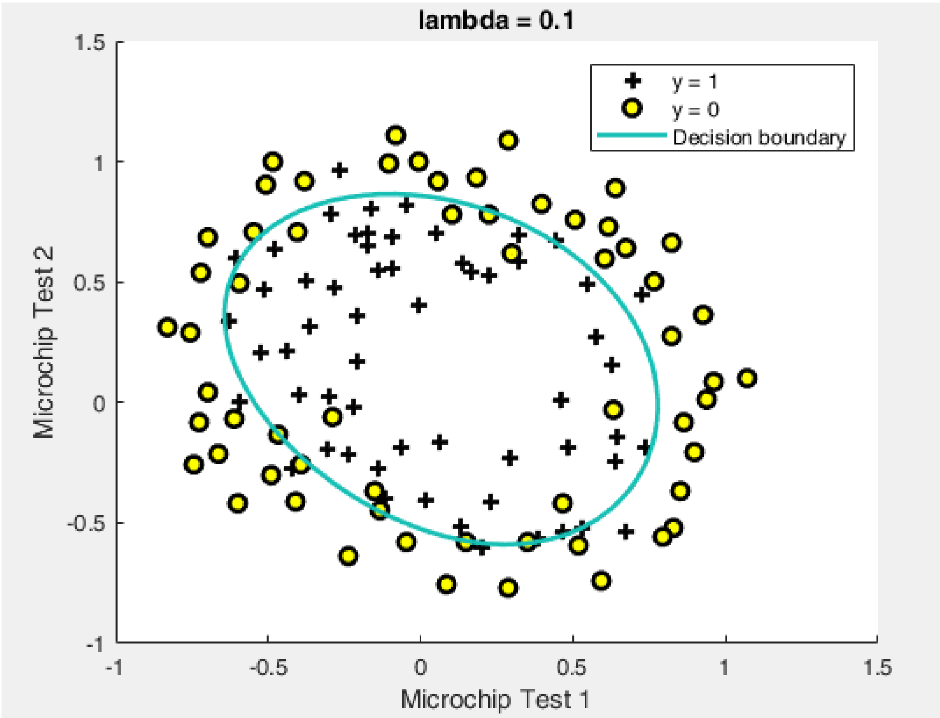
Labor_04

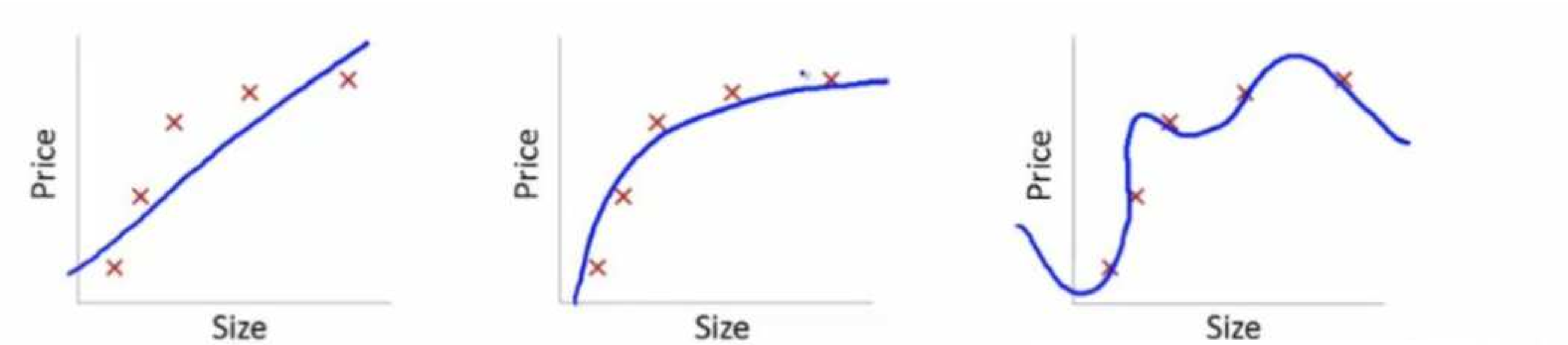
LOGISTIC REGRESSION NON LINEAR CASE

Non Linear Boundary



Polynomial features

$x_1 \; x_2 \;\Rightarrow\; 1 \; x_1 \; x_2 \; x_1^2 \; x_1x_2 \; x_2^2 \; x_1^3 \; x_1^2x_2 \; x_1x_2^2 \; x_2^3$

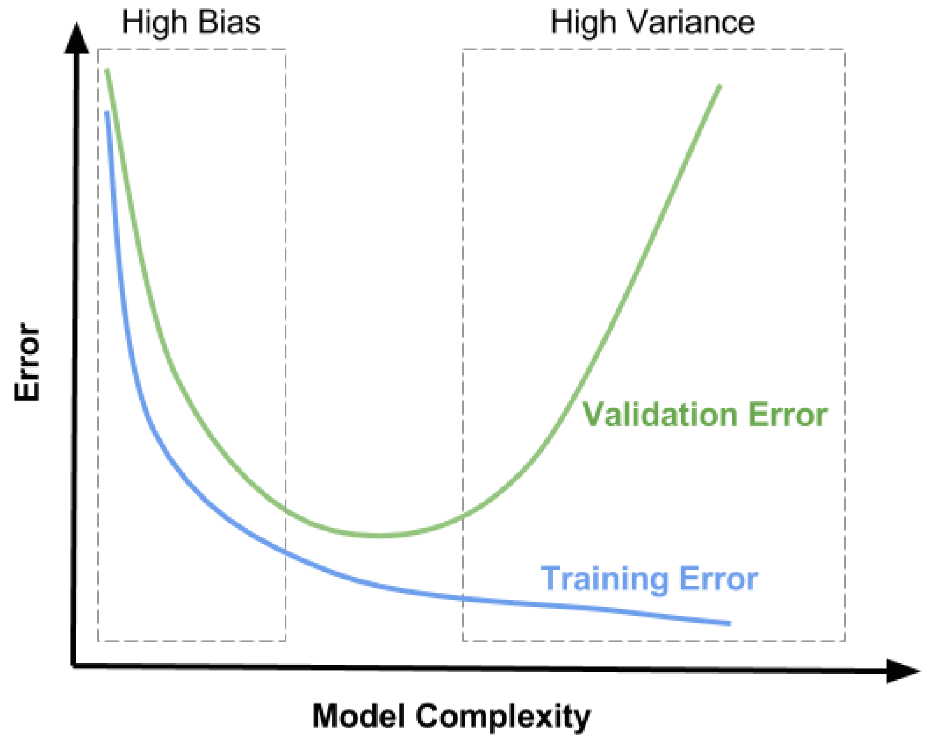


$w_0 + w_1x$
"Underfit"
"High Bias"

$w_0 + w_1x + w_2x^2$
Just Right

$w_0 + w_1x + w_2x^2 + w_3x^3 + w_4x^4$
"Overfit"
"High Variance"

Bias vs Variance



Problem

- Not enough feature: underfit

- Too many features: overfit

Solution:

- Reduce number of features
  - Manually
  - Model selection
- Regularization
  - Keep all the features, but reduce magnitude/value of weight

DO NOT PENALIZE THE BIAS!!!

Regularization

$$C(w) = \frac{1}{2m} \sum_{i=1}^{m} (h_w(x^i) - y^i)^2 + \lambda \sum_{j=1}^{n} w_j^2$$

If λ large: algorithm result in underfitting
(fails to fit even the training set)

Regularized Logistic Regression

Regularized Logistic Regression:

Repeat{

$$w_0 := w_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_w(x^i) - y^i) \cdot x_0^i$$

$$w_j := w_j - \alpha[\frac{1}{m} \sum_{i=1}^{m} (h_w(x^i) - y^i) \cdot x_j^i + \frac{\lambda}{m} w_j$$

}

Cost function and derivative

$$C(w) = [-\frac{1}{m}\sum_{i=1}^{m} y^i \cdot log(h_w(x^i)) + (1-y^i)\cdot log(1-h_w(x^i))] + \frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2$$

$$\frac{\partial}{\partial w_0}C(w) = \frac{1}{m}\sum_{i=1}^{m}(h_w(x^i) - y^i)\cdot x_0^i + 0$$

$$\frac{\partial}{\partial w_j}C(w) = \frac{1}{m}\sum_{i=1}^{m}(h_w(x^i) - y^i)\cdot x_j^i + \frac{\lambda}{m}w_j$$

Gradient Descent:

Want min $min_W\{C(W)\}$

Algorithm:

repeat until convergence {

$W_j := W_j - \mu\frac{\partial}{\partial W_j}C(W)$

}

Zipped in: fmincg() or fminunc() function